COMPARATIVE ANALYSIS OF EXPONENTIAL SMOOTHING MODELS FOR FORECASTING THE WATER QUALITY INDEX

H. Dissanayake*[*] *and R. Punchi-Manage
Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

Surface water quality plays a significant role in maintaining healthy ecosystems and supporting both human and aquatic life, making reliable monitoring and forecasting essential due to increasing pollution and environmental changes. The objective of this study is to forecast the Weighted Arithmetic Water Quality Index (WAWQI) at 20 sampling sites along the River Thames using three time series models: Single Exponential Smoothing (SES), Holt's (Double) Exponential Smoothing (HES), and Holt-Winters (Triple) Exponential Smoothing (HWES), and to evaluate and compare the forecasting performance using three accuracy metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to identify the most suitable model for sites with varying temporal patterns (stable, trending, or seasonal). Water quality data collected from 20 sampling sites along the River Thames between March 2009 and September 2017 were arranged chronologically. Each model was applied after linearly interpolating missing data and splitting the series into training and testing sets in an 80:20 ratio. The results revealed that model performance varied depending on the temporal patterns of WQI data. SES performed well at sites with stable conditions, such as TC8, TC12, TC13, and TC17. TC8 recorded the lowest RMSE (1.49), MAE (1.12), and MAPE (1.68%), indicating high forecasting accuracy. The HES model accounts for both the level and trend components of time series data and generally does not outperform the SES model for most sites, since most of the sites lack trend components. At TC20, the HES model showed the highest accuracy, with RMSE of 2.11, MAE of 1.38, and MAPE of 1.96%. HWES achieved the best performance across the majority of monitoring sites, particularly those exhibiting clear seasonal patterns in WQI fluctuations. In contrast, volatile sites (TC15) resulted in higher forecast errors (MAPE >15%) regardless of the model applied. These findings suggest that model selection should consider the underlying temporal characteristics of WQI behavior at each site: HWES for seasonal patterns, HES for trending series, and SES for stable conditions. These insights can aid water management authorities in proactive pollution control and sustainable resource planning by enabling accurate water quality forecasting.

Keywords: Water Quality Index, Single Exponential Smoothing, Holt's Exponential Smoothing, Holt-Winters Exponential Smoothing

[*] *Corresponding Author: hashinid@sci.pdn.ac.lk*

COMPARATIVE ANALYSIS OF EXPONENTIAL SMOOTHING MODELS FOR FORECASTING THE WATER QUALITY INDEX

H. Dissanayake* ***and R. Punchi-Manage***
Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

INTRODUCTION

Surface water quality is vital for ecosystem health and human use, requiring reliable monitoring amid rising pollution and environmental change. This study analyzes water quality in the River Thames using the Weighted Arithmetic Water Quality Index (WAWQI), which integrates multiple chemical parameters into a single metric. The primary objective is to forecast WAWQI using Single Exponential Smoothing (SES), Holt's (Double) Exponential Smoothing (HES), and Holt-Winters (Triple) Exponential Smoothing (HWES), and assess their accuracy via RMSE, MAE, and MAPE. These models apply unequal weights to past data, making them suitable for capturing environmental trends (Hyndman *et al.*, 2024). By comparing these methods, the study seeks to identify the most effective approach under varying hydrological conditions to support informed water management and sustainability efforts.

METHODOLOGY

Study Area

This study examines the River Thames in Southern England (51°38'N–51°83'N, 01°75'E–00°56'E) using data from 20 sites collected between March 2009 and September 2017. Physical and chemical parameters were obtained from the Environmental Information Data Centre (EIDC), managed by the UK Centre for Ecology & Hydrology.

Weighted Arithmetic Water Quality Index

The Water Quality Index (WQI) simplifies complex chemical data into a single value to assess surface water quality. This study uses the WA WQI to classify water quality by purity level, following a multi-step calculation process. First, a unit weight (W_n) was assigned to each water quality parameter using the formula: $W_n = k/S_n$ where S_n represents the standard permissible value for the nth WQ parameter and k is the constant of proportionality, calculated as $k = 1/\Sigma[1/S_n]$. The standard values (S_n) were based on the WHO-recommended standards. Then, the quality rating scale (Q_n) for each parameter was calculated using the formula: $Q_n = [(V_n - V_{id})/(S_n - V_{id})] \times 100$, where, V_n is the estimated value of nth WQ parameter and V_{id} is the ideal value of nth parameter in pure water (V_{id} for pH = 7 and zero for all other variables). Finally, the WAWQI for each observation was calculated using $WQI = \Sigma Q_n W_n / \Sigma W_n$.

Single Exponential Smoothing (SES)

To forecast the WQI over time for each monitoring site, Single Exponential Smoothing was used. The SES model forecasts future values using a weighted average of past observations, with more recent observations receiving higher weights. It has the form, $\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$ where $\hat{Y}_{t+1}$ is the forecasted value of the series at time $t+1$, Y_t is the observed value at time t, and α is the smoothing constant, where $0 < \alpha < 1$ (Chatfield, 2000).

Holt's (Double) Exponential Smoothing (HES)

This method is suitable for time series data exhibiting trends, as it estimates both the level and the trend components of the series. The model is defined as: $L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$, $T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$ where, L_t is the local level at time t, α is the smoothing constant, T_t is the local estimate of the growth rate at time t, and β is the smoothing constant for the trend. The h steps ahead forecast is $\hat{Y}_{t+h} = L_t + hT_t$ (Chatfield, 2000).

Holt-Winters (Triple) Exponential Smoothing (HWES)

To model and forecast temporal patterns in Water Quality Index (WQI) values at each sampling site, the Holt-Winters Exponential Smoothing method with additive seasonality was employed. This technique is suited for time series data that exhibit trend and seasonal patterns. The additive model is defined as: $L_t = \alpha(Y_t - I_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1})$, $T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$, $I_t = \delta(Y_t - L_t) + (1-\delta)I_{t-s}$ where, level, L_t , the growth rate, T_t, and the seasonal index, I_t at time t. δ is the smoothing constant for the seasonal index, and s is the seasonal period. The h steps ahead forecast is $\hat{Y}_{t+h} = L_t + hT_t + I_{t-s+h}$ (Chatfield, 2000).

The data were arranged chronologically, and missing values were linearly interpolated. Time series diagnostics were applied to each site to identify temporal patterns. Many sites, including TC1, TC10, TC4, etc., exhibited seasonality, and TC20 showed an upward trend. Sites such as TC12, TC13, & TC17 remained relatively stable. For all three methods, the data for each site were split into training and testing sets using an 80:20 ratio. Model performance was assessed using three standard evaluation metrics: RMSE, MAE, and MAPE. These were calculated on the test dataset for each site to evaluate forecast accuracy. The implementation was carried out in R.

RESULTS AND DISCUSSION

Single Exponential Smoothing (SES)

SES model performance varied across the 20 sites. Smoothing parameter α ranged between 0.23 and 0.76 across sites, with lower values indicating stronger smoothing and a better fit for stable series. RMSE ranged from 1.49 (TC8) to 10.07 (TC2), MAE from 1.12 to 9.02, and MAPE from 1.67% to 22.77% (Table 1). TC8 showed the highest accuracy with consistently low errors. RMSE reflects average forecasting error, with values under 10% indicating reliable predictions. MAE, less sensitive to outliers, offers a clear measure of average error. MAPE expresses error as a percentage of actual WQI values, with <10% indicating high accuracy and >20% suggesting issues like non-stationarity or abrupt quality changes. Figure 1 shows SES fits and forecasts for each River Thames site. For low-MAPE sites, fitted values tracked WQI trends well, capturing short-term fluctuations. In contrast, higher MAPE sites showed greater forecast deviation, revealing SES's limitations with sudden shifts or seasonal changes, and suggesting the need for models that account for trend or seasonality.

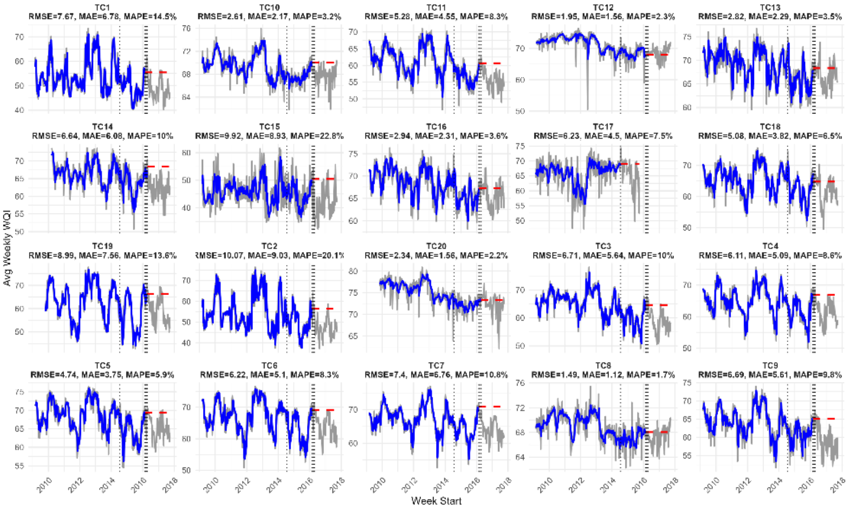

Figure 1: SES fitted models and forecasted WQI values for each monitoring site of the River *Thames* (The grey line indicates the actual values, and the blue line represents the fitted values during the training period. The red line indicates the SES forecasts).

Holt's (Double) Exponential Smoothing (HES)

The application of HES for forecasting WQI across multiple monitoring sites revealed important insights into both the model's predictive capabilities and the underlying trends in water quality data. The α values ranged from 0.23 to 0.78, while the trend smoothing parameters β were 0.0001 for all sites. The HES model accounts for both the level and trend components of time series data, and generally does not outperform the SES model for most sites, since most of the sites lack trend components. Figure 2 illustrates the HES fitted models and forecasted WQI values for each monitoring site of the River *Thames*. RMSE values ranged from 1.48 (TC8) to 10.09 (TC15), MAE varied from 1.13 (TC8) to 9.09 (TC15), and MAPE ranged from 1.70% (TC8) to 23.19% (TC15) (Table 1).

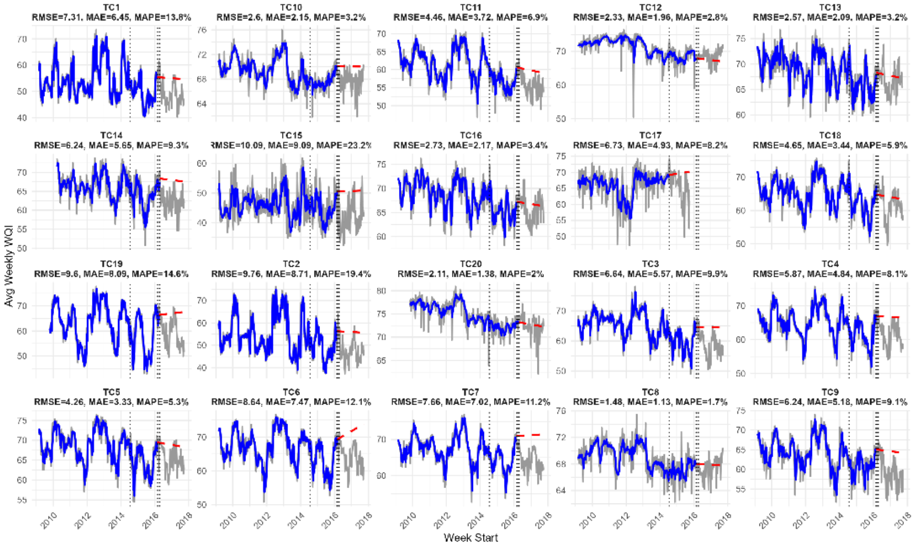

Figure 2: HES fitted models and forecasted WQI values for each monitoring site.

Holt-Winters (Triple) Exponential Smoothing (HWES)

To capture the underlying trends and seasonal patterns in the WQI across the monitoring sites, Holt-Winters triple exponential smoothing was employed using the additive seasonal model.

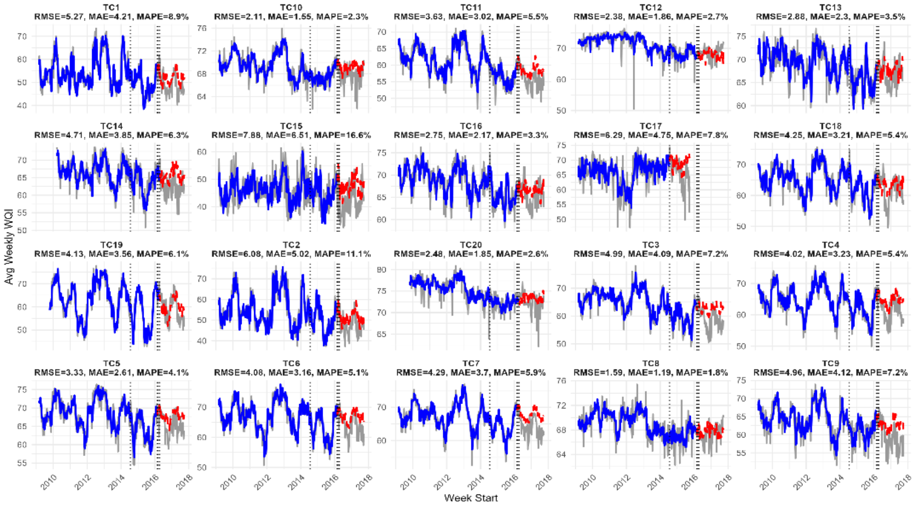

Figure 3: H-WES fitted models and forecasted WQI values for each monitoring site.

Figure 3 presents Holt-Winters (HWES) fitted models and WQI forecasts for each River Thames site, showing smooth, repeating seasonal patterns typical of the method. At sites like TC2 and TC19, forecasts closely matched actual values, while at TC17, and TC13, the model struggled with sudden fluctuations, producing flatter predictions. RMSE ranged from 1.59 (TC8) to 7.88 (TC15), MAE from 1.19 to 6.51, and MAPE from 1.78% to 16.16% (Table 1). HWES performed best where seasonality was strong, while SES outperformed in sites with weak or no seasonal patterns.

Table 1: Performance matrices for each site. (grey color indicates minimum values)

SITE	Single			Double (Holt)			Triple (Holt-Winter)			*n*
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	
TC1	7.67	6.78	14.47	7.31	6.45	13.75	5.27	4.21	8.89	403
TC10	2.61	2.17	3.24	2.60	2.15	3.22	2.11	1.55	2.32	403
TC11	5.28	4.55	8.34	4.46	3.72	6.85	3.63	3.02	5.52	404
TC12	1.95	1.56	2.26	2.33	1.96	2.82	2.38	1.86	2.70	394
TC13	2.82	2.29	3.48	2.57	2.09	3.16	2.88	2.30	3.47	393
TC14	6.64	6.08	9.97	6.24	5.65	9.27	4.71	3.85	6.33	351
TC15	9.92	8.93	22.77	10.09	9.09	23.19	7.88	6.51	16.61	395
TC16	2.94	2.31	3.59	2.73	2.17	3.36	2.75	2.17	3.34	395
TC17	6.23	4.50	7.47	6.73	4.93	8.17	6.29	4.75	7.79	298
TC18	5.08	3.82	6.54	4.65	3.44	5.88	4.25	3.21	5.42	395
TC19	8.99	7.56	13.63	9.60	8.09	14.58	4.13	3.56	6.11	376
TC2	10.07	9.03	20.09	9.76	8.71	19.40	6.08	5.02	11.11	403
TC20	2.34	1.56	2.23	2.11	1.38	1.96	2.48	1.85	2.62	370
TC3	6.71	5.64	9.98	6.64	5.57	9.87	4.99	4.09	7.23	402
TC4	6.11	5.09	8.55	5.87	4.84	8.14	4.02	3.23	5.42	402
TC5	4.74	3.75	5.91	4.26	3.33	5.25	3.33	2.61	4.11	400
TC6	6.22	5.10	8.30	8.64	7.47	12.06	4.08	3.16	5.14	403
TC7	7.40	6.76	10.77	7.66	7.02	11.19	4.29	3.70	5.91	402
TC8	1.49	1.12	1.68	1.48	1.13	1.70	1.59	1.19	1.78	404
TC9	6.69	5.61	9.85	6.24	5.18	9.11	4.96	4.12	7.20	404

CONCLUSIONS/RECOMMENDATIONS

This study aims to compare and forecast the performance of SES, HES, and HWES models for predicting the WAWQI across 20 sites along the River Thames. Model performance was evaluated via RMSE, MAE, and MAPE. By analyzing sites with stable, trending, and seasonal patterns, this study identifies conditions under which each method is most suitable. HWES yielded the lowest errors at most sites, identified as the most suitable for sites with seasonal patterns. SES performed well on stable data, while HES suited trend-dominated sites. Volatile sites resulted in higher errors across all methods. Method selection should match site-specific patterns: HWES for seasonal, HES for trending, and SES for stable conditions. These findings indicate the significance of determining a site-specific model, which allows for more precise water quality forecasting and supports proactive pollution control and sustainable resource planning actions by water management authorities.

REFERENCES

Chatfield, C. (2000). *Time-Series Forecasting*. Boca Raton: Chapman & Hall/CRC.

Hamner, B., & Frasco, M. (2018). *Metrics: Evaluation Metrics for Machine Learning* (R package version 0.1.4).

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Hyndman, R., Petropoulos, F., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Razbash, S., Wang, E., & Yasmeen, F. (2024). *forecast: Forecasting functions for time series and linear model* (R package version 8.23.0).